



Article

The Reliability and Medical Students' Appreciation of Certainty-Based Marking

Špela Smrkolj ^{1,2,*} , Enja Bančov ¹ and Vladimir Smrkolj ¹

¹ Faculty of Medicine, University of Ljubljana, 1000 Ljubljana, Slovenia; enja.bancov12@gmail.com (E.B.); vladimir.smrkolj@gmail.com (V.S.)

² Department of Obstetrics and Gynaecology, University Medical Centre, 1000 Ljubljana, Slovenia

* Correspondence: spela.smrkolj@mf.uni-lj.si; Tel.: +386-31-865-432

Abstract: Certainty-Based Marking (CBM) involves asking students not only the answer to an objective question, but also how certain they are that their answer is correct. In a mixed method design employing an embedded approach with a quasi-experimental design, we have examined the use of CBM during a 5-week Gynaecology and Obstetrics course. The study was conducted as a non-mandatory revision exam with two additional questionnaires on Moodle. Majority of students perceive CBM as fair (78%) and useful (94%). Most students would immediately want CBM to be used for revision exams, but more practice would be needed for CBM to be used in graded exams. The lowest self-evaluation of knowledge was mostly seen by worst (less than 70% Accuracy) and best achievers (more than 90% Accuracy); the worst achievers probably have knowledge gaps, and the best achievers probably correctly guessed at least one question. Our findings conclude that CBM does not discriminate any learner type ($p = 0.932$) and does not change the general distribution of the exam scores, since there is no significant differences between Certainty-Based Score ($M = 80.4\%$, $SD = 10.4\%$) and Accuracy ($M = 79.8\%$, $SD = 11.1\%$); $t(176) = 0.8327$, $p = 0.4061$. These findings are widely applicable, as learner type study models are used extensively in education. In the future, larger samples should be studied and the implementation of CBM on question types other than MCQ should be investigated.

Keywords: certainty-based marking; confidence-based learning; undergraduate medical education; online exam; self-evaluation



Citation: Smrkolj, Š.; Bančov, E.; Smrkolj, V. The Reliability and Medical Students' Appreciation of Certainty-Based Marking. *Int. J. Environ. Res. Public Health* **2022**, *19*, 1706. <https://doi.org/10.3390/ijerph19031706>

Academic Editor: Paul B. Tchounwou

Received: 14 December 2021

Accepted: 30 January 2022

Published: 2 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The differentiation between knowledge and luck on exams has been puzzling researchers for many decades. Some of the older studies, such as Coombs et al. [1], Hevner [2] and Soderquist [3] already tried to incorporate certainty or confidence into exams. Gardner-Medwind has extensively incorporated CBM into revision exams at University College London, where he had over 9000 sessions of CBM use. He optimised the certainty bonus factors and the scoring of certainty [4]. CBM upgrades Multiple Choice Questions (MCQ), so that students must indicate how certain they are in their answer. The more reliable is their self-evaluation of their knowledge, and the greater reward they receive (CBM Bonus) [5,6]. When using CBM, one must indicate the degree of certainty in the given answer; the scale is not defined exactly, but many authors, including Gardner-Medwind, use a 3-point scale [6]. The descriptions of the scale are *low-mid-high*, which correspond to 50%-67%-80% certainty. The scoring then follows according to a scoring table, in which high certainty receives a maximum award or penalty, and low certainty has no penalty if wrong, but receives the least points if the answer is correct. There are then multiple indicators which describe a student's performance on the exam; the percentage of correct answers (Accuracy), overall CBM Bonus (CBM Bonus) and the Certainty-Based Score. The CBM Bonus (given in %) shows a student's deviation from the standard curve in terms of the self-evaluation of certainty, while the Certainty-Based Score is the sum of Accuracy and CBM Bonus. Automatic

analysis of the answers at each certainty level can also give students feedback, whether they have been over or under confident at each certainty level [6,7]. Considering all of the above, the use of CBM should enable both the examiner and the student to more precisely evaluate the student's knowledge [8].

However, the integration of CBM in the curriculum is only present in a small number of faculties, which widens the discrepancy between students' and teachers' appreciation of the use of CBM and its actual availability in the curriculum [9]. To yield results, CBM must not be integrated solely as a type of examination, but rather as a part of Confidence-based learning, which is a philosophical idea—that knowledge is not only the correctness of answers but also the person's certainty. That is manifested as the difference between "I know" and "I think" [10]; a difference most important in medicine, as an error can result in tragedy. In fact, medical workers are encouraged to consult in case of doubt, which can only be possible if a person can discriminate between true knowledge and uncertain knowledge. The above-described philosophy should be taught to students throughout the entire study program, as it indirectly changes the student's study behaviour towards a performance improvement. This indirect effect is thought to result from the student's self-regulated learning—the student will isolate the topics where she or he lacks knowledge and focus on them [11,12]. Our experience with medical students is that they are aware of the fact that their knowledge is incomplete but they find it hard to quantify and isolate the topics where they have the least knowledge. Moreover, it is common that they have conceptual misunderstandings; consequently, they are not aware of the fact that they lack knowledge [13–15]. To support this, Hendriks et al. [9] concluded that CBM is greatly appreciated by students; however, more accessible options to integrate CBM should be developed, which has now been done [16]. Regular revision exams can be used to enhance medical education, as they promote better retention of information. This phenomenon is known as the testing effect [17]. Furthermore, the phenomenon can be further expanded by providing more detailed feedback, although this is yet to be confirmed by solid evidence [18]. Again CBM is promising a self-assessment tool to regularly give students the ability to test their knowledge and receive more detailed feedback. However, the feedback received from CBM is not as detailed as the one given by a teacher [15,19–22]. In contrast, computer based learning methods show no overt effect when compared to traditional study methods [23]. Moreover, the factor of luck in MCQ is a known design problem and when not enough care is taken to actively discourage the random guessing, exams become unreliable [24–29]. CBM was designed to tackle the issues presented above—identifying students' conceptual misunderstandings, identifying students' areas for improvement and actively discouraging random guessing [30–32].

It should be noted that the students' performance on all exams may differ due to their adaption to the type of the exam. This could be connected to the student's learner style, and effort should be made to not discriminate against either learner style if learner style models are concluded to be valid. There are many learning style models and schemas described in the literature; however, they have been followed with controversy and questionable validity. One systematic review tried to evaluate the 14 most influential models, whereas they listed over 70 models that have been developed [33]. One of the most commonly used models is the VARK model, developed by N. Fleming and C. Mills in 1992 [34]. It differentiates between four basic learner types (modalities): Visual, Auditory, Read/Write and Kinesthetic. Despite the VARK model's debated validity it is used widely between students, teachers and researchers [35–38]. Between different question types, MCQ was shown to be least sensitive to learner style [39]. For CBM to be useful in the curriculum it must maintain this trait.

In this article we are trying to tackle the implementation stage of the CBM addition to the curriculum. We are evaluating the reliability of CBM in terms of student learning types and their self-reported study efforts; moreover, we aim to gather their feedback on implementing the CBM in the curriculum, to deduce if students would like to have CBM used in exams. Considering the above, we have set the following research questions: Is

CBM reliable in terms of students' learner types? and Do students appreciate the use of CBM in curriculum?

2. Method

2.1. Study Design

The mixed method design employed was an embedded approach with a quasi-experimental design. The quantitative method was quasi-experimental between-subjects approach utilizing non-manipulated independent variables based on a closed-ended pretest questionnaire. Qualitative data were collected as a post-test questionnaire [40–42]. Independent variables were learner type, number of times the student rehearsed the topic, when the student started learning and the student's self-evaluated knowledge level, whereas dependent variables were Accuracy, CBM bonus and Certainty-Based Score. To sum up, the study was conducted as a non-mandatory revision exam with two additional questionnaires: one before (quantitative data) and one after the exam (quantitative and qualitative data).

2.2. Data Collection

Both the CBM exam and the questionnaires were built on the Moodle open-source learning platform [16]. We have chosen Moodle, as it is a free learning platform, which is accessible to any faculty, moreover it readily supports CBM exams as a free plug-in. Before the study was conducted, the students were given a short description of the study, were told what CBM is and how to use it. Firstly, a questionnaire about the learner style and study efforts was given to the students. The learner style part of the first questionnaire was a translated VARK questionnaire [43], while the study efforts part asked the students to determine when they started learning, how many times they rehearsed the subject matter and how they perceive their knowledge of the subject. As a base for the learner style questionnaire we have used an English VARK questionnaire version 8.01 [44]. The VARK questionnaire consists of 16 MCQ, where each answer is characteristic for one of the four modalities of VARK theory. The student's modality is determined by the highest number of answers that correspond to that modality. We have had the questionnaire translated by a C1 certified English speaker from English into Slovene. The translated questionnaire was independently back-translated. Minor changes to the initial translation were made after the back translation. The validity of the questionnaire was approved by expert committee, which consisted of three members. The first member was an expert on the topic of interest and the methodology; the second was the forward translator and the third was the backward translator. The translated questionnaire was then pilot tested with a small group of students ($n = 32$) as recommended by Tsang et al. [45]. After students in our study finished the first questionnaire, the CBM exam was unlocked and students were given 20 min to solve 20 MCQ questions with CBM and only one correct answer. After the CBM exam they were not immediately shown their achievement, but they answered the second questionnaire. The second questionnaire asked them about the difficulties they had when using CBM, whether their achievement was impaired because of CBM and if they see the use of CBM as fair. In addition they had an option to leave general feedback in the form of a text-box field. All together the duration of the study was 60 min. The translated questionnaire is available in the Supplementary Materials.

2.3. Sample

We have examined the use of CBM during a 5-week Gynaecology and Obstetrics course at the Faculty of Medicine at the University of Ljubljana. The students who participated in the study were fifth-year medical students ($n = 89$) from four different rotations (groups). The mean age of students was 23.04 years ($SD = 0.49$ year). Three quarters of students were females (74%), the rest were males (26%). They all consented to participate and provide data for the study.

2.4. Data Analysis

Student performance was compared based on their questionnaire response using one-way ANOVA, where more than three categories per variable were present, and using the t-test, where only two categories per variable were present. Only the students who completed the questionnaires before and after the CBM exam were included in the analysis of the effect of learner type and students' self reported study efforts on the CBM performance. In the analysis of the distribution of scores between Accuracy and Certainty-Based Score, the students who did not complete the questionnaires were also included. Quantitative data from the second survey (feedback) were analysed with basic descriptive statistics. Qualitative data were organized and examined. To aid in the analysis of the qualitative data we have developed some tags (arguments that have occurred in the feedback) and tried to extract the most important ones. Statistical analysis was performed in OriginPro 2021 software.

3. Results

3.1. Participation

The participation rate in the study was 62% (89 out of 143 students). Whoever did not want to participate in the study was given the opportunity to solve the revision exam without CBM or with CBM, while their data were not included in the study. Out of 89 students who participated in the study, 45% (40 students) did not solve the questionnaires before and after the CBM revision exam.

3.2. Student Previous Experience

Students' responses to the questionnaire before the CBM exam are presented in Figure 1. Surprisingly, 29% of students had no previous experience with online exams, moreover only one student was previously familiar with CBM.

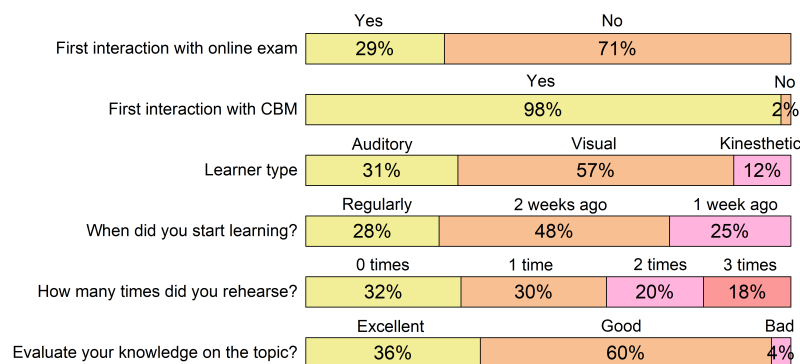


Figure 1. Students' responses to the questionnaire before the CBM exam ($n = 49$). The labels on the left represent the question or a series of questions that were given in the questionnaire. The text above the stacked bars represents the possible answers, while the width of the bar represents the portion of students who selected that answer. The data for *Learner type* questions has been analysed and only the determined proportions for each learner type are represented.

Most students were of the visual learner type, followed by auditory and kinesthetic learner types. Moreover, the ratio of learner types was similar in all the groups that we included in the study. A quarter of students said they learned regularly, while about a half started learning two weeks before the study. The date of the study had been known in advance as well as the dates of all the other study activities and the topics that were on the CBM revision exam.

3.3. Students' Feedback

The majority of students indicated that they perceive CBM as fair (78%) and useful (94%) as seen in Figure 2.

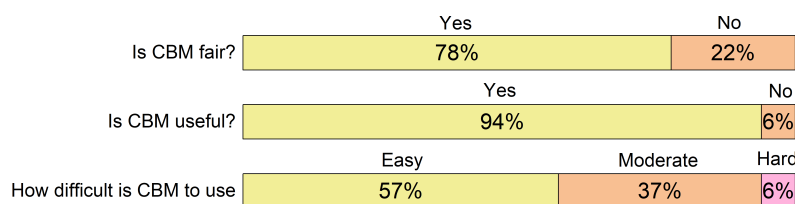


Figure 2. Students’ responses to the questionnaire after the CBM exam ($n = 49$). The labels on the left represent the question or a series of questions that were given in the questionnaire. The text above the stacked bars represents the possible answers, while the width of the bar represents the portion of students who selected that answer.

Twenty-five students also described their experience in the text box field in the questionnaire. From the feedback we received, the most popular concern is that someone who showed less absolute knowledge (Accuracy) can receive a better grade, some however thought that CBM is useful, but it cannot be directly implemented, as it requires a mental shift. Three students were concerned that the CBM Bonus is too big, and it should be reduced, their argument is that currently overly self-critical students would be negatively discriminated. One student pointed out that the questions must be stated clearly to avoid ambiguity, as that could result in students being less certain in their answers. Several students reported that including CBM in non-revision exams would increase their stress-level at these exams. Students were generally satisfied with CBM experience and have stated it would be very beneficial to be used at least as a revision tool. Some (15 students) reported they had been able to identify knowledge gaps using the automatic feedback that was given at the end of CBM (after the second questionnaire). All of the data from the open type questions (text box) are available upon enquiry.

3.4. Reliability of CBM in Terms of Learner Types and Students’ Self-Reported Study Efforts

There is no statistically significant link between students’ learner types and their achievement in any of the CBM parameters (Accuracy, CBM Bonus, Certainty-Based Score), as seen in Table 1.

Table 1. Results of mean-comparison tests between different student’s qualities (Learner type, Learning start, Rehearsal times, Self-evaluated knowledge) and performance indicators (Accuracy, CBM Bonus and Certainty-Based Score).

Independent Variable	p-Value		
	Accuracy	CBM Bonus	Certainty-Based Score
Learner type	0.900	0.711	0.932
Learning start	0.512	0.940	0.453
Rehearsal times	0.593	0.862	0.423
Self-evaluated knowledge	0.901	0.492	0.718

However, when comparing students’ answer accuracy with their average CBM mark, the deviation from the standard curve is clearly seen (Figure 3). In general, the students’ achievements are distributed equally above and below the standard curve; however, top-achievers in terms of accuracy (above 90%) have mostly scored below the standard curve. Similarly, the bottom-achievers in terms of accuracy (below 70%) have scored generally below the standard curve. The best achievers in terms of CBM mark are those in the middle range of accuracy (70–90%).

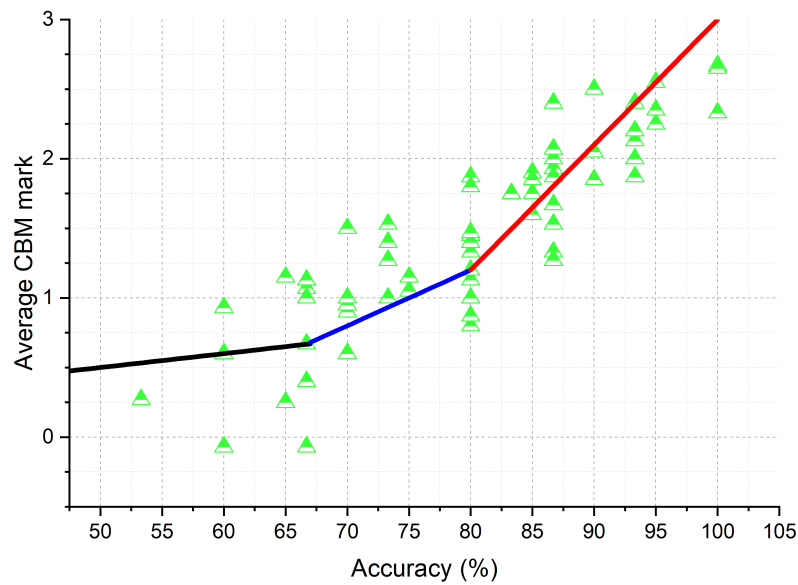


Figure 3. Average CBM mark (=All points/number of questions) versus Accuracy shows the deviation from the standard curve, which represents random guessing.

Comparing the grade distribution of Accuracy and Certainty-Based Score (Accuracy + CBM Bonus) in (Figure 4) yields that there is no significant differences between Certainty-Based Score ($M = 80.4\%$, $SD = 10.4\%$) and Accuracy ($M = 79.8\%$, $SD = 11.1\%$); $t(176) = 0.8327$, $p = 0.4061$.

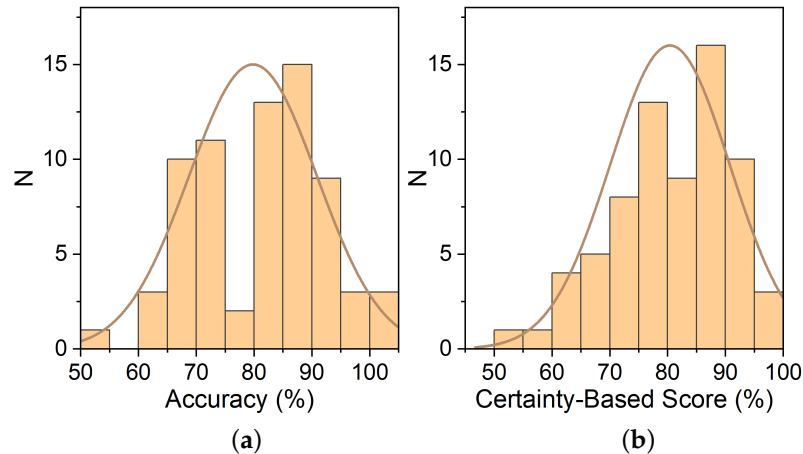


Figure 4. (a) Distribution of Accuracy scores. (b) Distribution of Certainty-Based Score.

However, the Certainty-Based Score has more negative skewness (offset to the right) than Accuracy, which suggests that the grades shifted to the right; however, both the minimum and the maximum grade are lower in terms of Certainty-Based Score than in terms of Accuracy. Although the Certainty-Based Score distribution seems to be shifted to the right, only 60% of students received positive CBM Bonus. Consequently, it is obvious that the distribution has been roughly preserved; however, the students themselves have changed their relative order in many cases.

4. Discussion

Increased use of computer-based multiple choice questioning in the COVID pandemic has enabled us to upgrade content using CBM. The pioneering work of Gardner-Medwin and colleagues [1–6,30] has been exploited to support diverse educational goals. As CBM

has never been incorporated in the curriculum at our Faculty, we have considered it is high time we researched the reliability and medical students' appreciation of CBM.

4.1. Students' Previous Experience

Most of the students who participated in the study were already familiar with Moodle as an online study and exam platform, which suggests that there was some use of online learning tools already before the study. However still 29% of students were new to online exams. This highlights that there is a substantial inequality between students in terms of experience with technology, which must be reduced to minimum, before these online tools are used in curriculum [46]. It must be noted that our study was conducted during the first year of COVID-19 pandemic, therefore the first group in the study was new to online studying, while later groups were more accustomed to it. Similarly, this was discussed by Aguilera-Hermida [47] where she found an increased use of online educational technology during the pandemic by almost 45%. Students' increased experience with online study tools is beneficial for the inclusion of CBM into the curriculum. During the pandemic, online learning tools have had a crucial role in filling the gaps and ensuring uninterrupted education, as was discussed by Chatterjee and Chakraborty [48]. Therefore, it is our duty to change the study programs in a way to be more future proof, which certainly includes online learning platforms, some of which already include CBM, as has been demonstrated in our study with an example from Moodle.

4.2. Students' Feedback

The feedback we received from students has shown that the use of CBM could potentially be successful and well received if enough preparation would have been done. Most importantly, the implementation must be done gradually for students who are not familiar with the concept, as was mentioned by few students in their feedback. It is a reasonable request from students, as only one student in our study had previous experience with CBM. Therefore it is not a surprise that many students reported in their feedback that they only think CBM should be used for revision exams. Almost all students (94%) saw CBM as a valuable tool for revision, which is also one of the key aspects that the original author of CBM, Gardner-Medwin, has promoted [6,30]. In contrast, Schoendorfer and Emmett [49] reported that only around 60% of students appreciated the benefits of CBM, moreover 67% of students in their study have reported that it took them more time to solve exam and consequently 30% of students saw CBM as a "waste of time thinking about certainty". Similarly, Hendriks et al. [9] reported that 32.7% of students were against the use of CBM. Perhaps the difference between our students' high appreciation of CBM compared to the previously discussed studies is due to our study being conducted during pandemic, when students were deprived of many other forms of feedback. Considering what students in our study have written in the feedback, we can observe that many of them felt positive about the auto-generated feedback about their over or under-confidence. While students were encouraged to reflect on their knowledge throughout the process of solving the CBM exam, the benefits of the process would have been lesser if they had not received the feedback per certainty level and topic. Per their responses, it is only with that information that allows them to improve their knowledge where it gaps. Furthermore, Hendriks et al. [9] reported similar conclusions—52.2% of students in their study considered auto-generated feedback useful. These observations answered our research question, which asked if students appreciated CBM.

We have anticipated that most of the students rehearsed the study matter at least once, but it was unexpected that 18% of them rehearsed it three times. Because the duration of the course was five weeks, and the study was performed during the third week of the course, it is highly doubtful that they have rehearsed all the topics three times. More so since it had been made clear that this study would not affect their course grade in any way and that rather than studying hard for the CBM exam, they should not change their study plan and take the CBM exam as a revision exam. This is definitely a weakness of our

study, which could be tackled with a larger sample. Most (60%) of our students evaluated their knowledge to be good, 4% evaluated it to be bad and 36% evaluated it to be excellent. From the feedback we received, we concluded that we should devise the self-evaluation of knowledge per topic, as it is easier to evaluate it in that way, rather than questioning about the general knowledge, which is hard to quickly and accurately self-evaluate.

4.3. Reliability of CBM in Terms of Learner Types and Students' Self-Reported Study Efforts

We have not established a statistically significant difference between the achievements of students with different learner types. In the light of our first research question we can conclude that CBM is reliable in terms of students' learner types. Although there is no consensus about whether learner types are indeed valid, its use is unquestionably widespread, which makes our finding important [35–38]. We can also confirm that the findings from Holley and Jenkins [39], which suggest that the MCQs have high reliability in terms of learner types. Furthermore we found that this trait is preserved when using MCQ in conjunction with CBM.

As we anticipated, there is no statistical difference between the achievements of students with or without the use of CBM. This is due to Gardner-Medwin [4]'s analysis of several thousand CBM revision exams and his optimization of the standard curve and the bonus factor, to yield good psychometric reliability and decrease the inflation of grades (reduces the predictive value of the exam). As mentioned in the Introduction, the CBM should not be implemented as an isolated tool, but rather as a part of Confidence-based learning philosophy taught through the entire program. In this case students would be more familiar with the concept, its benefits and the platform, as suggested by the following authors Gardner-Medwin [6], Agrawal et al. [11] and Larsen et al. [17]. Otherwise, the exam results would not be reliable as some students who are less agile with the concept of CBM would be negatively discriminated against. Further study should be conducted to determine how much practice is needed by the students to be able to solve exams without being hindered by the complexity of the exam platform or the CBM.

Surprisingly, we have not established a statistical difference in achievements of students between the number of times they rehearsed or when they started learning. We can only speculate about the cause of that, perhaps we should ask students to evaluate their knowledge at each topic, then their self-evaluations could be more objective. In further studies, more participants should be included; consequently, the analysis of their achievements in a group will be more reliable, yielding results that could be—if CBM could differentiate students with more profound knowledge—better than just the accuracy of answers. However, it is already visible from the results in Figures 3 and 4 that while the distribution of scores between Accuracy and Certainty-based Score is relatively unchanged, the students' in-group order must have been changed. That means that the use of CBM most likely will not alter the general score distribution of a group, as that is a function of the difficulty of questions and general knowledge in a group, but will rather reorder the students in-group to correct for their good or bad confidence, similarly noted by Hendriks et al. [9]. This is important because CBM cannot be used to replace good and valid questions in exams; CBM can only enhance their discrimination power. We have observed that the worst self-evaluators of knowledge were the worst and best-achievers. The best achievers probably at least partially scored highly due to luck; however, the bottom achievers were probably unfamiliar with the topics and perhaps had certain knowledge misconceptions. Wu et al. [22] and Gardner-Medwin [4] observed the same phenomenon that the best self-evaluators of knowledge were mid-achieving students.

5. Conclusions

This study explored the use, appreciation and reliability of CBM with medical students. CBM is not a new tool, but its benefits have not been fully utilised yet. Moreover, it is easily setup with widely used learning platforms such as Moodle. Our findings showed that most students perceive CBM as a positive addition to their study, but some are still sceptical

about whether CBM can be more than just a revision tool. Adding certainty ratings to MCQ answers seems to reinforce formative testing with better feedback. It also offers deeper insight into the successful delivery of online course content, identifying areas for improvement of teaching and content delivery as well as exam question design. Secondly, our findings conclude that CBM does not discriminate against any learner type (based on VARK questionnaire) and it does not change the general distribution of the exam scores. In fact, our findings are relevant to the whole educational community, since the use of learner type models is widespread. In future studies, a larger sample should be studied and it would also be wise to examine the implementation of CBM on question types other than MCQ.

5.1. Usefulness of CBM

CBM in a formative assessment improves student course appreciation, presumably by helping students to evaluate their mastery level and identify misconceptions. Certainty-based learning inclusion in formative assessments may impinge on exam designs and course evaluations, hence representing an important and useful assessment tool for teaching staff as well [5,9–11]. During the COVID pandemic, the study of medicine has been largely distance-based, with the exception of clinical exercises. For this reason, it is especially important that students have a tool such as CBM available in the learning process, so that they can gain more reliable knowledge, which they can then put into daily practice in a clinical setting. Moreover we must encourage students to aid their studies with internet based methods, consequently future lock-downs will be less stressful for them and much easier to overcome [50]. Perhaps most important is the enhancement of learning with certainty-based marking self-tests [6]. It is argued that such self-tests should be available to students to explore in private, making mistakes and finding strengths and weaknesses without submitting any data to the view of teachers.

5.2. Limitations

Using a larger sample size, we could reduce random errors, due to dishonesty in answers to the questionnaires. Moreover our study did not take into account the possibility of academic dishonesty such as cheating on the revision exam. This is a design flaw that is inherent due to the nature of the observations. To explain better: we evaluated online exams which students solved from home, and if it involved a supervision of some kind it would make the study biased. Some students also did not want to participate in the study. If there is some underlying reason as to why they did not want to participate, such as their low self-confidence, that would make our sample unrepresentative and our findings less valid. Moreover, students expressed their perception in the middle of pandemics. In such conditions, students could have increased fear or stress, which could negatively affect some of the students' performances.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijerph19031706/s1>, Table S1: Translated VARK questionnaire (in Slovene).

Author Contributions: Conceptualization, V.S. and Š.S.; methodology, V.S. and E.B.; software, V.S.; formal analysis, V.S.; investigation, V.S., Š.S. and E.B.; resources, V.S., Š.S. and E.B.; data curation, V.S.; writing—original draft preparation, V.S. and Š.S.; writing—review and editing, V.S., Š.S. and E.B.; visualization, V.S.; project administration, Š.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data that supports the findings of this study are available from the corresponding author, Š.S., upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CBM Certainty-Based Marking
MCQ Multiple choice question

References

1. Coombs, C.; Milholland, J.E.; Womer, F.B. The assessment of partial knowledge. *Educ. Psychol. Meas.* **1956**, *16*, 13–37. [CrossRef]
2. Hevner, K. A method of correcting for guessing in true-false tests and empirical evidence in support of it. *J. Soc. Psychol.* **1932**, *3*, 359–362. [CrossRef]
3. Soderquist, H.O. A New Method Weight. Scores A True-False Test. *J. Educ. Res.* **1936**, *30*, 290–292. [CrossRef]
4. Gardner-Medwin, T. Optimisation of Certainty-Based Assessment Scores. 2013. Available online: https://tmedwin.net/~ucgbarg/tea/IUPS_2013a.pdf (accessed on 10 November 2021).
5. Ahlgren, A. Reliability, Predictive Validity, and Personality Bias of Confidence Weighted Scores. 1969. Available online: <https://www.ucl.ac.uk/lapt/ahlgren.pdf> (accessed on 10 November 2021).
6. Gardner-Medwin, A. Certainty-Based Marking: Rewarding Good Judgment of What Is or Is Not Reliable. 2008. Available online: <https://www.tmedwin.net/~ucgbarg/tea/innovation2008.html> (accessed on 10 November 2021).
7. Luetsch, K.; Burrows, J. Certainty rating in pre-and post-tests of study modules in an online clinical pharmacy course—A pilot study to evaluate teaching and learning. *BMC Med. Educ.* **2016**, *16*, 267. [CrossRef] [PubMed]
8. Barr, D.A.; Burke, J.R. Using confidence-based marking in a laboratory setting: A tool for student self-assessment and learning. *J. Chiropr. Educ.* **2013**, *27*, 21–26. [CrossRef]
9. Hendriks, W.J.A.J.; Bakker, N.; Pluk, H.; de Brouwer, A.; Wieringa, B.; Cambi, A.; Zegers, M.; Wansink, D.G.; Leunissen, R.; Klaren, P.H.M. Certainty-based marking in a formative assessment improves student course appreciation but not summative examination scores. *BMC Med. Educ.* **2019**, *19*, 178. [CrossRef]
10. Hunt, D.P. The concept of knowledge and how to measure it. *J. Intellect. Cap.* **2003**, *4*, 100–113. [CrossRef]
11. Agrawal, S.; Norman, G.R.; Eva, K.W. Influences on medical students' self-regulated learning after test completion. *Med. Educ.* **2012**, *46*, 326–335. [CrossRef]
12. Ende, J. Feedback in Clinical Medical Education. *JAMA* **1983**, *250*, 777–781. [CrossRef]
13. Bienstock, J.L.; Katz, N.T.; Cox, S.M.; Hueppchen, N.; Erickson, S.; Puscheck, E.E. To the point: Medical education reviews—Providing feedback. *Am. J. Obstet. Gynecol.* **2007**, *196*, 508–513. [CrossRef] [PubMed]
14. Bing-You, R.; Hayes, V.; Varaklis, K.; Trowbridge, R.; Kemp, H.; McKelvy, D. Feedback for Learners in Medical Education: What Is Known? A Scoping Review. *Acad. Med.* **2017**, *92*, 1346–1354. [CrossRef] [PubMed]
15. Wigton, R.S.; Patil, K.D.; Hoellerich, V.L. The effect of feedback in learning clinical diagnosis. *J. Med. Educ.* **1986**, *61*, 816–822. [CrossRef]
16. Moodle. 2020. Available online: https://docs.moodle.org/311/en/Using_certainty-based_marking (accessed on 1 September 2021).
17. Larsen, D.P.; Butler, A.C.; Roediger, H.L., III. Test-enhanced learning in medical education. *Med. Educ.* **2008**, *42*, 959–966. [CrossRef]
18. Colthart, I.; Bagnall, G.; Evans, A.; Allbutt, H.; Haig, A.; Illing, J.; McKinstry, B. The effectiveness of self-assessment on the identification of learner needs, learner activity, and impact on clinical practice: BEME Guide no. 10. *Med. Teach.* **2008**, *30*, 124–145. [CrossRef] [PubMed]
19. Clark, C. The Impact of Confidence-Based Marking on Unit Exam Achievement in a High School Physical Science Course. In *Graduate Research Papers*; University of Northern Iowa: Cedar Falls, IA, USA, 2020.
20. Nožica, B.; Čarapina, M.; Staničić, O. Certainty based marking: The use of modified grading schema in multiple choice quizzes. In Proceedings of the 6th International Conference of Education, Research and Innovation, Seville, Spain, 18–20 November 2013; pp. 2160–2165.
21. Cisar, S.M.; Cisar, P.; Pinter, R. True/false questions analysis using computerized Certainty-Based Marking tests. In Proceedings of the 2009 7th International Symposium on Intelligent Systems and Informatics, Subotica, Serbia, 25–26 September 2009; pp. 171–174. [CrossRef]
22. Wu, Q.; Vanerum, M.; Agten, A.; Christiansen, A.; Vandenabeele, F.; Rigo, J.M.; Janssen, R. Certainty-Based Marking on Multiple-Choice Items: Psychometrics Meets Decision Theory. *Psychometrika* **2021**, *86*, 518–543. [CrossRef] [PubMed]
23. Cook, D.A.; Levinson, A.J.; Garside, S.; Dupras, D.M.; Erwin, P.J.; Montori, V.M. Internet-based learning in the health professions: A meta-analysis. *JAMA* **2008**, *300*, 1181–1196. [CrossRef]
24. Burton, R.F. Quantifying the Effects of Chance in Multiple Choice and True/False Tests: Question selection and guessing of answers. *Assess. Eval. High. Educ.* **2001**, *26*, 41–50. [CrossRef]
25. Frary, R.B. Formula Scoring of Multiple-Choice Tests (Correction for Guessing). *Educ. Meas. Issues Pract.* **1988**, *7*, 33–38. [CrossRef]

26. Lesage, E.; Valcke, M.; Sabbe, E. Scoring methods for multiple choice assessment in higher education—Is it still a matter of number right scoring or negative marking? *Stud. Educ. Eval.* **2013**, *39*, 188–193. [[CrossRef](#)]
27. Wilkinson, K.; Dafoulas, G.; Garelick, H.; Huyck, C. Are quiz-games an effective revision tool in Anatomical Sciences for Higher Education and what do students think of them? *Br. J. Educ. Technol.* **2020**, *51*, 761–777. [[CrossRef](#)]
28. Wu, Q.; De Laet, T.; Janssen, R. Elimination Scoring Versus Correction for Guessing: A Simulation Study. In *Quantitative Psychology*; Wiberg, M., Culpepper, S., Janssen, R., González, J., Molenaar, D., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 183–193.
29. Bradbard, D.A.; Parker, D.F.; Stone, G.L. An Alternate Multiple-Choice Scoring Procedure in a Macroeconomics Course. *Decis. Sci. J. Innov. Educ.* **2004**, *2*, 11–26. [[CrossRef](#)]
30. Gardner-Medwin, T.; Curtin, N. Certainty-Based Marking (CBM) for Reflective Learning and Proper Knowledge Assessment. 2007. Available online: https://www.reap.ac.uk/reap/reap07/Portals/2/CSL/t2%20-%20great%20designs%20for%20assessment/raising%20students%20meta-cognition/Certainty_based_marking_for_reflective_learning_and_knowledge_assessment.pdf (accessed on 1 September 2021).
31. Hassmén, P.; Hunt, D.P. Human Self-Assessment in Multiple-Choice Testing. *J. Educ. Meas.* **1994**, *31*, 149–160. [[CrossRef](#)]
32. Leclercq, D. Confidence marking: Its use in testing. *Eval. Educ.* **1982**, *6*, 161–287. [[CrossRef](#)]
33. Coffield, F.; Moseley, D.; Hall, E.; Ecclestone, K. *Learning Styles and Pedagogy in Post-16 Learning: A Systematic and Critical Review*; Learning and Skills Research Centre: London, UK, 2004.
34. Fleming, N.D.; Mills, C. Not Another Inventory, Rather a Catalyst for Reflection. *Improv. Acad.* **1992**, *11*, 137–155. [[CrossRef](#)]
35. Hernandez, J.; Vasan, N. Learning styles among medical students: Kinesthetic learners approach to learning anatomy. *FASEB J.* **2018**, *32*, 508. [[CrossRef](#)]
36. Husmann, P.R.; O’Loughlin, V.D. Another Nail in the Coffin for Learning Styles? Disparities among Undergraduate Anatomy Students’ Study Strategies, Class Performance, and Reported VARK Learning Styles. *Am. Assoc. Anat.* **2019**, *12*, 6–19. [[CrossRef](#)] [[PubMed](#)]
37. Hernandez, J.E.; Vasan, N.; Huff, S.; Melovitz-Vasan, C. Learning Styles/Preferences Among Medical Students: Kinesthetic Learner’s Multimodal Approach to Learning Anatomy. *Med. Sci. Educ.* **2020**, *30*, 1633–1638. [[CrossRef](#)]
38. Pashler, H.; McDaniel, M.; Rohrer, D.; Bjork, R. Learning Styles: Concepts and Evidence. *Psychol. Sci. Public Interest* **2008**, *9*, 105–119. [[CrossRef](#)]
39. Holley, J.H.; Jenkins, E.K. The Relationship Between Student Learning Style and Performance on Various Test Question Formats. *J. Educ. Bus.* **1993**, *68*, 301–308. [[CrossRef](#)]
40. Price, P.C.; Jhangiani, R.; Chiang, I.C.A.; Leighton, D.C.; Cuttler, C. Setting Up a Factorial Experiment. Available online: <https://opentext.wsu.edu/carriecuttler/chapter/9-1-setting-up-a-factorial-experiment/> (accessed on 10 November 2021).
41. Light, G.; Calkins, S.; Luna, M.; Drane, D. Assessing the Impact of a Year-Long Faculty Development Program on Faculty Approaches to Teaching. *Int. J. Teach. Learn. High. Educ.* **2009**, *20*, 168–181.
42. Creswell, J.W. *Planning, Conducting, and Evaluating Quantitative and Qualitative Research*; Pearson: Boston, MA, USA, 2012; pp. 534–576.
43. Hadi, P.; Jamil, S.; Javaher, K.; Masood, Y.; Mohammad, R.; Yasemi, M.Y.; Nahal, M.M.; Karim, H. Using VARK Approach for Assessing Preferred Learning Styles of First Year Medical Sciences Students: A Survey from Iran. *J. Clin. Diagn. Res.* **2014**, *8*, GC01–GC04. [[CrossRef](#)]
44. Limited, V.L. The VARK Questionnaire (Version 8.01). 2019. Available online: <https://vark-learn.com/the-vark-questionnaire/> (accessed on 2 March 2020).
45. Tsang, S.; Royse, C.; Terkawi, A. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi J. Anaesth.* **2017**, *11*, S80–S89_203_17. [[CrossRef](#)] [[PubMed](#)]
46. Mukhtar, K.; Javed, K.; Arooj, M.; Sethi, A. Advantages, Limitations and Recommendations for online learning during COVID-19 pandemic era. *Pak. J. Med. Sci.* **2020**, *36*, S27–S31. [[CrossRef](#)] [[PubMed](#)]
47. Aguilera-Hermida, A.P. College students’ use and acceptance of emergency online learning due to COVID-19. *Int. J. Educ. Res. Open* **2020**, *1*, 100011. [[CrossRef](#)]
48. Chatterjee, I.; Chakraborty, P. Use of Information Communication Technology by Medical Educators Amid COVID-19 Pandemic and Beyond. *J. Educ. Technol. Syst.* **2021**, *49*, 310–324. [[CrossRef](#)]
49. Schoendorfer, N.; Emmett, D. Use of certainty-based marking in a second-year medical student cohort: A pilot study. *Adv. Med. Educ. Pract.* **2012**, *3*, 139–143. [[CrossRef](#)]
50. Almendingen, K.; Morseth, M.S.; Gjølstad, E.; Brevik, A.; Tørris, C. Student’s experiences with online teaching following COVID-19 lockdown: A mixed methods explorative study. *PLoS ONE* **2021**, *16*, e0250378. [[CrossRef](#)]